caBIG cancer Biomedical Informatics Grid

NCICB's Pathway Interaction Database

Carl F. Schaefer, Joshua Phillips
National Cancer Institute
Center for Bioinformatics

August 3, 2004

NATIONAL CANCER INSTITUTE

Center for Bioinformatics

Agenda

- Motivation
- Representation
- New caBIO-like API
- Long term plans

Motivation: Examples of Questions (1 of 2)

- What downstream interactions could be affected, directly or indirectly, by a mutation in a particular protein or by a change in the abundance of a particular protein?
- How many parallel, independent paths are known to lead to the same event (e.g. activation of a particular protein)?
- What anomalies (mutation, over-expression, under-expression) might theoretically result in a failure of the DNA repair mechanism? Might these same anomalies disrupt other processes?

Motivation: Examples of Questions (2 of 2)

- Loss of heterozygosity of a region within 17q21 has been detected in 30% of primary breast tumors. What cellular processes would be most directly affected by a loss of function of genes in this region? Are any candidate genes in the region closely connected to each other in pathway networks?

- … questions about cause/effect networks

Representation (1 of 2)

- Pathway: directed graph
  - node: molecule or event or condition
  - edge: role of molecule/condition in an event
  - interaction: event & its connected molecules/conditions
- Molecule type:
  - protein | complex | compound | rna
  - or families (e.g. EC_2.7.7.15 includes PCYT1A, PCYT1B)
- Event type:
  - reaction | modification | transcription | translocation
  - or any GO BP type

Representation (2 of 2)

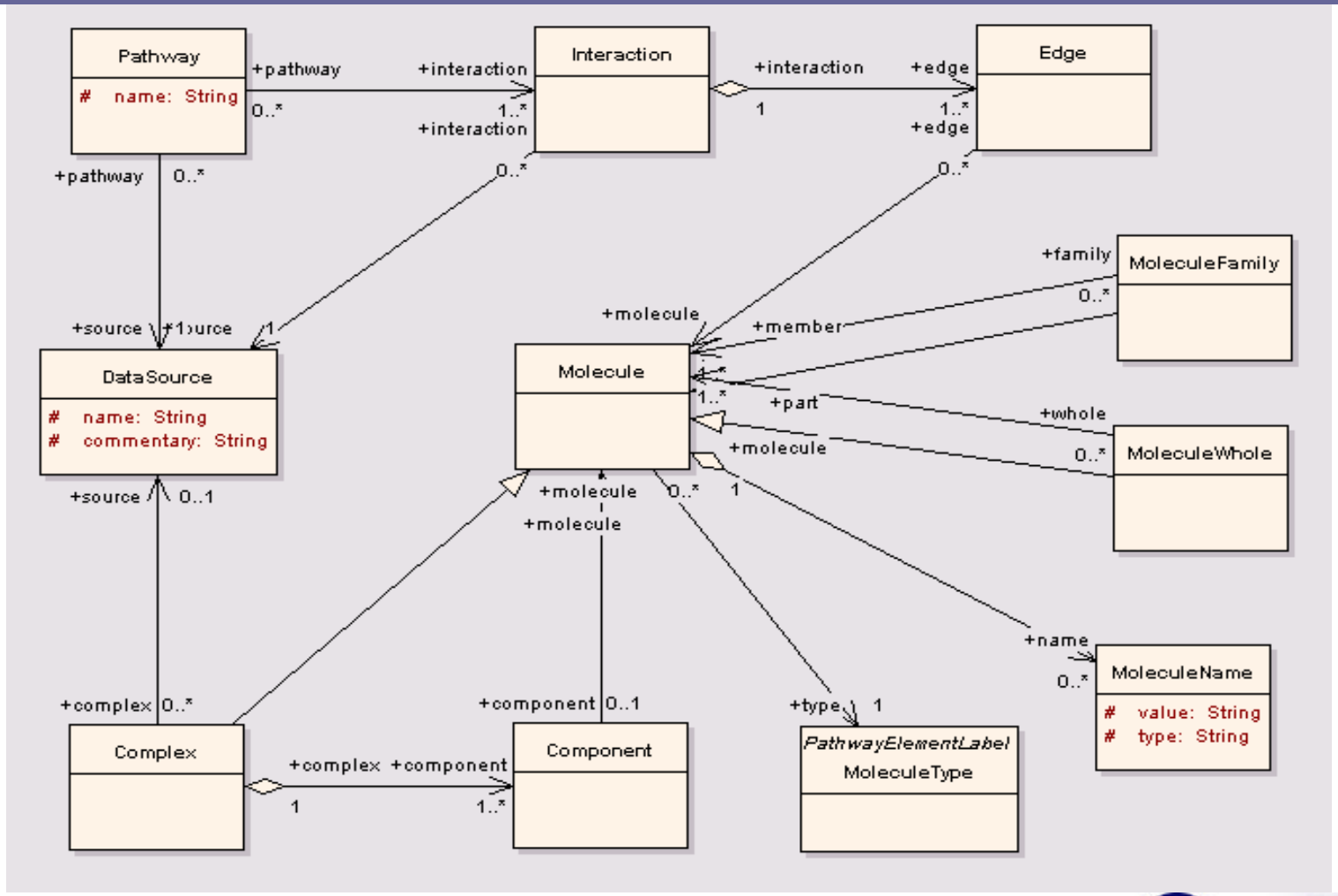
- Condition type:
  - any GO BP type
- Role type:
  - input | output | agent | inhibitor
  - or any GO MF type
- Molecule location
  - any GO CC type
  - specified at point of use
- Posttranslational modification
  - abstract terms (e.g. “active”)
  - specified at point of use

Current Contents (all human)

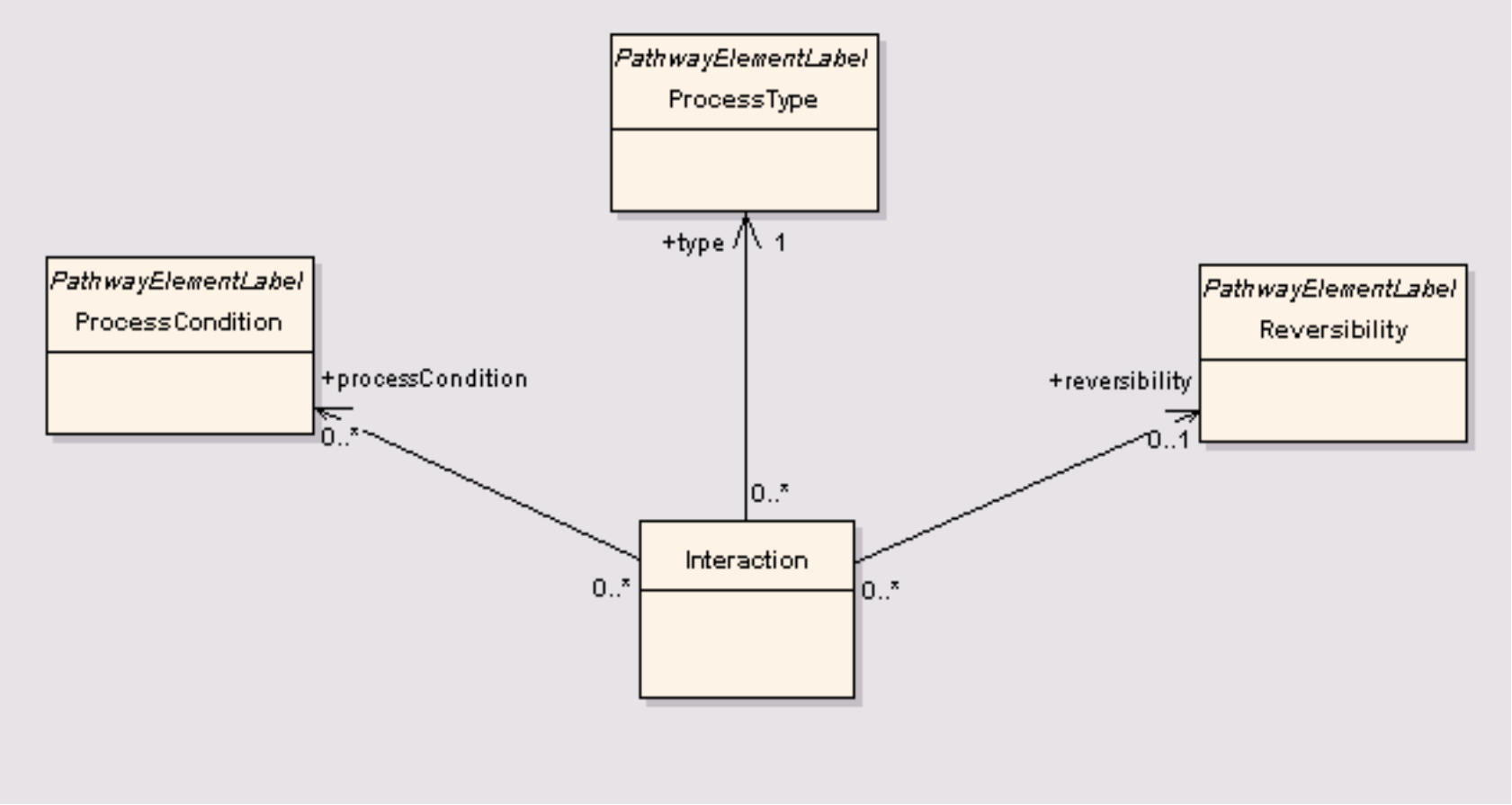
	BioCarta	KEGG
Pathways	259	85
Interactions	3064	4207

http://cmap.nci.nih.gov/PW

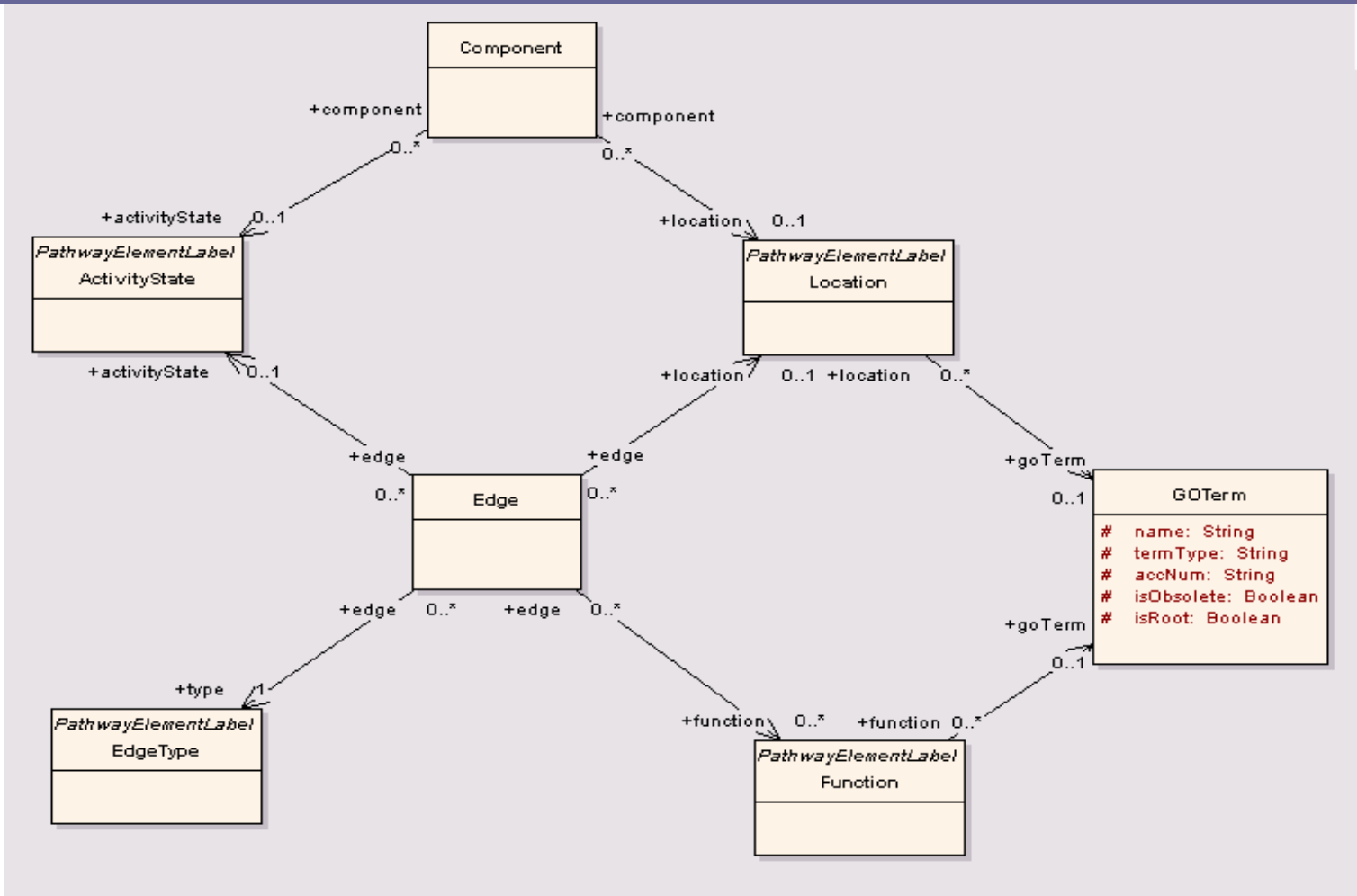
Primary Domain Model



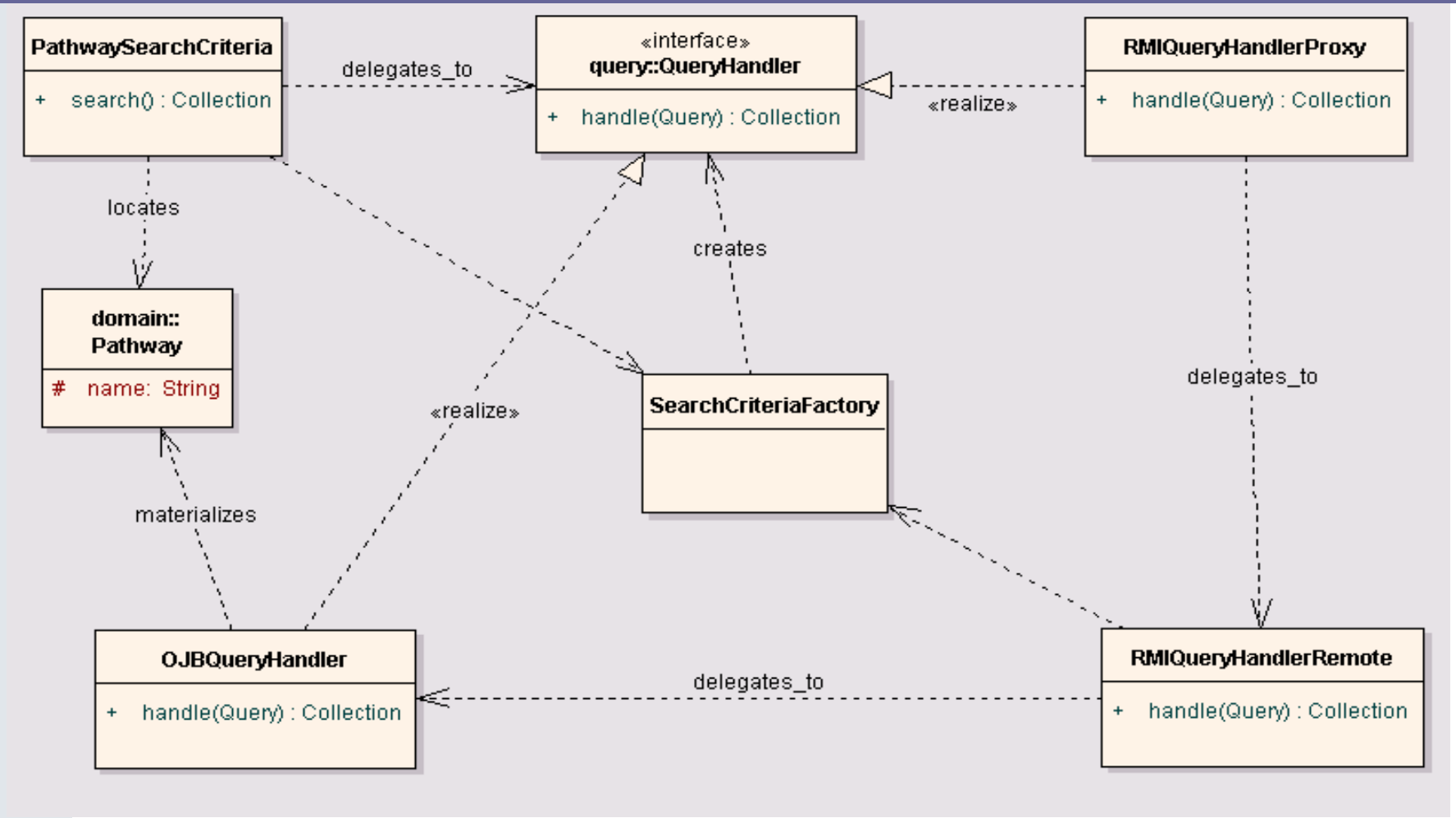
Interaction Labels



Edge & Component Labels



Remote Architecture



Example

- Retrieve all existing pathways that involve these molecules: VEGF, CAV1, GUCYB12. Then print out Interaction information for each.

Code Snippet

```
PathwaySearchCriteria psc = (PathwaySearchCriteria)
    SearchCriteriaFactory.newSearchCriteria("Pathway");
InteractionSearchCriteria isc = ...
EdgeSearchCriteria esc = ...
MoleculeSearchCriteria msc = ...
MoleculeNameSearchCriteria mnsc = ...
Collection names = new ArrayList();
names.add("VEGF");
names.add("CAV1");
names.add("GUCYB12");
mnsc.setValue(names, BooleanOperationEnum.AND, false);
msc.setNameCriteria(mnsc);
esc.setMoleculeCriteria(msc);
isc.setEdgeCriteria(esc);
psc.setInteractionCriteria(isc);
Collection pathways = psc.search();
```

Code Snippet

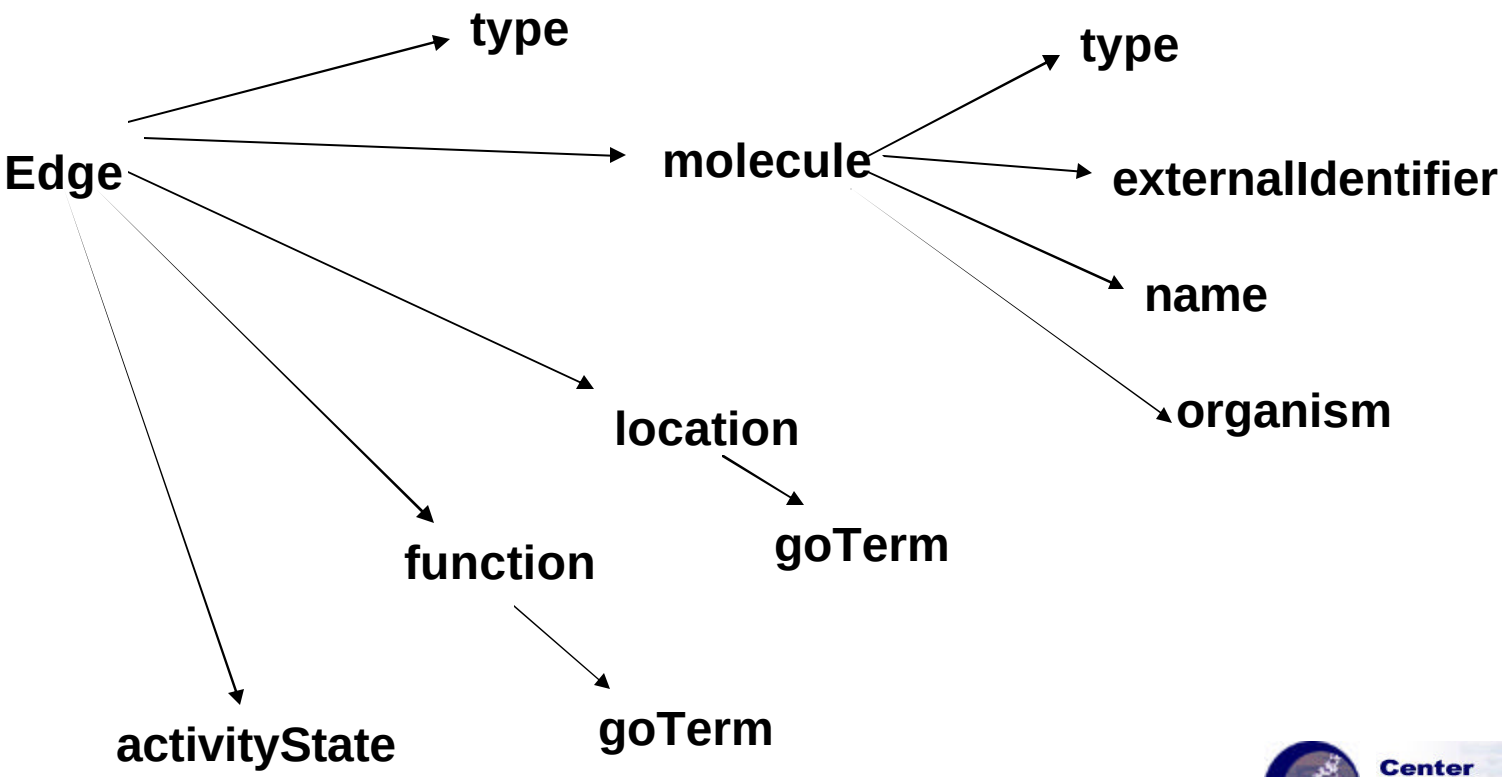
```
for(Iterator i = pathways.iterator(); i.hasNext();){
   Pathway p = (Pathway)i.next();
   System.out.println("Interactions for Pathway: " + p.getName());
   for(Iterator j = p.getInteraction().iterator(); j.hasNext();){
      Interaction inter = (Interaction)j.next();
      System.out.println("\ttype: " + inter.getType().getValue());
      for(Iterator k = inter.getProcessCondition().iterator(); k.hasNext();){
         ProcessCondition cond = (ProcessCondition)k.next();
         System.out.println("\tcondition: " + cond.getValue());
      }
   }
}
```

Output

```
Interactions for Pathway: Hypoxia-Inducible Factor in the
Cardivascular System
        type: transcription
        type: translocation
        type: translocation
        ...
Interactions for Pathway: Actions of Nitric Oxide in the Heart
        type: translocation
        type: modification
        type: modification
        ...
Interactions for Pathway: VEGF Hypoxia and Angiogenesis
        type: modification
        type: modification
        type: modification
        ...
```

Fetch Path

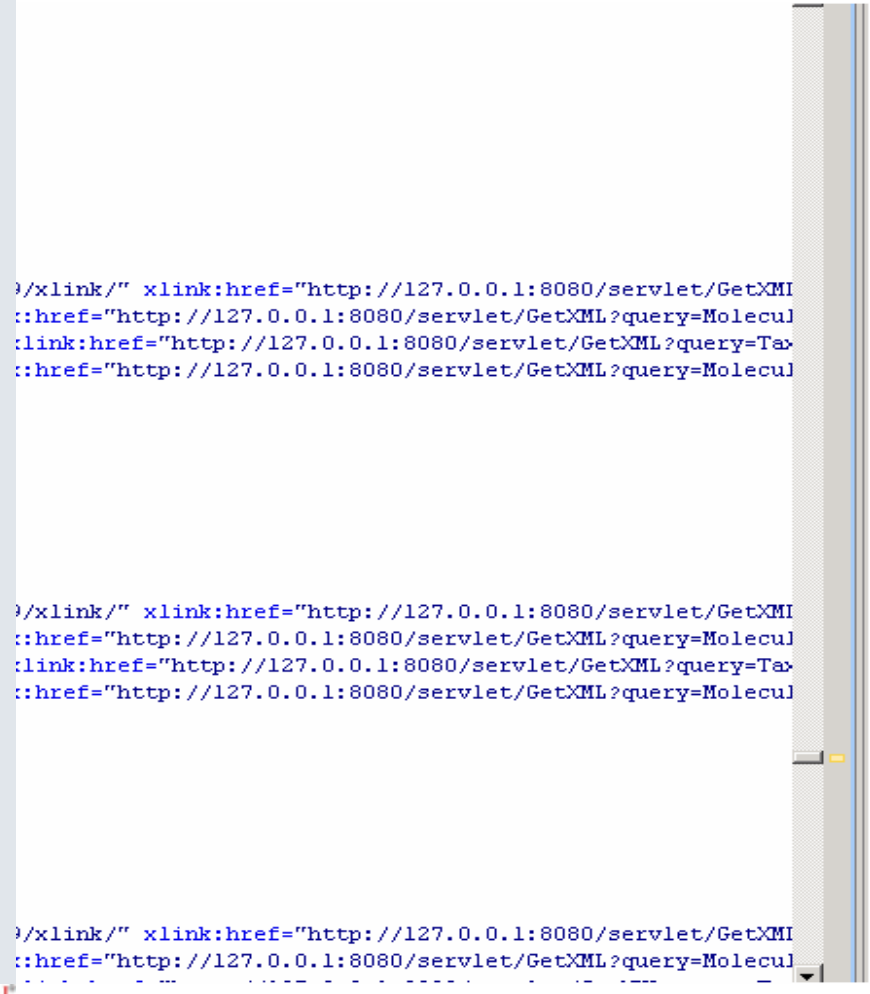
- The query API provides fine-grained control over object materialization.

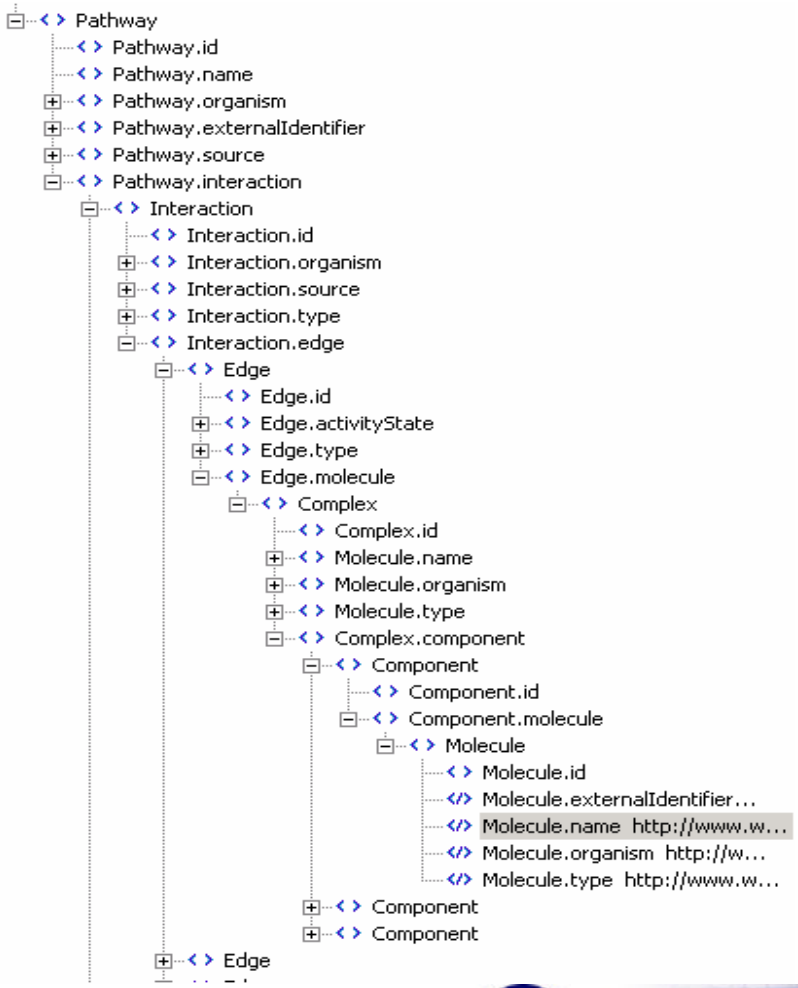


Applications

- Connect Molecules Query
  - Edge equality: molecule.id
- Group Interactions
  - Edge equality: activityState AND location AND molecule.id
- CMAP XML
- caBIO XML
- Dot Conversion

caBIO XML


/xlink/" xlink:href="http://127.0.0.1:8080/servlet/GetXMI
::href="http://127.0.0.1:8080/servlet/GetXML?query=Molecul
:link:href="http://127.0.0.1:8080/servlet/GetXML?query=Tax
::href="http://127.0.0.1:8080/servlet/GetXML?query=Molecul
/xlink/" xlink:href="http://127.0.0.1:8080/servlet/GetXMI
::href="http://127.0.0.1:8080/servlet/GetXML?query=Molecul
:link:href="http://127.0.0.1:8080/servlet/GetXML?query=Tax
::href="http://127.0.0.1:8080/servlet/GetXML?query=Molecul
/xlink/" xlink:href="http://127.0.0.1:8080/servlet/GetXMI
::href="http://127.0.0.1:8080/servlet/GetXML?query=Molecul


Pathway
Pathway.id
Pathway.name
Pathway.organism
Pathway.externalIdentifier
Pathway.source
Pathway.interaction
Interaction
Interaction.id
Interaction.organism
Interaction.source
Interaction.type
Interaction.edge
Edge
Edge.id
Edge.activityState
Edge.type
Edge.molecule
Complex
Complex.id
Molecule.name
Molecule.organism
Molecule.type
Complex.component
Component
Component.id
Component.molecule
Molecule
Molecule.id
Molecule.externalIdentifier...
Molecule.name http://www.w...
Molecule.organism http://w...
Molecule.type http://www.w...
Component
Component
Edge

CMAP XML

```
<NCICB_PID_XML>
    <Created />
    <Model>
        <MoleculeList>
            <Molecule id="20008">
                <Label label_type="molecule-type" value="co
                <ComplexComponent idref="20001">
                    <Label label_type="location" value="ext
                </ComplexComponent>
                <ComplexComponent idref="20004" />
                <ComplexComponent idref="20005" />
                <ComplexComponent idref="20006" />
                <ComplexComponent idref="20007" />
                <ComplexComponent idref="20003">
                    <Label label_type="location" value="pla
                </ComplexComponent>
            </Molecule>
            <Molecule id="20001">
                <Label label_type="molecule-type" value="pr
                <Name name_type="OF" value="TNF" />
                <Name name_type="AS" value="TNF-alpha" />
                <Name name_type="LL" value="7124" />
            </Molecule>
            <Molecule id="20004">
                <Label label_type="molecule-type" value="pr
                <Name name_type="OF" value="FADD" />
                <Name name_type="LL" value="8772" />
            </Molecule>
            <Molecule id="20005">
                <Label label_type="molecule-type" value="pr
                <Name name_type="OF" value="TRADD" />
                <Name name_type="LL" value="8717" />
            </Molecule>
            <Molecule id="20006">
                <Label label_type="molecule-type" value="pr
                <Name name_type="AS" value="RIP" />
                <Name name_type="OF" value="RIPK1" />
```

- NCICB_PID_XML
  - Created
  - Model
    - MoleculeList
    - InteractionList
      - Interaction 9595
        - Source
        - Label process-type
        - InteractionComponent 20008
        - InteractionComponent 20277
          - Label edge-type
          - Label activity-state
        - InteractionComponent 20277
      - Interaction 9605
      - Interaction 9602
      - Interaction 9592
      - Interaction 9606
      - Interaction 9596
      - Interaction 9603
      - Interaction 9593
      - Interaction 9599
      - Interaction 9597
      - Interaction 9609
      - Interaction 9598
      - Interaction 9608
      - Interaction 9600
      - Interaction 9594
      - Interaction 9604
      - Interaction 9591
      - Interaction 9607
    - PathwayList
      - Pathway 1104
        - Organism
        - Name
        - ShortName
        - Source
        - PathwayComponent 9595

Long Term Plans

- Data
  - Improved curation (exploring several possibilities)
  - New data sources: Reactome, BIND
- Data Model
  - Further elaboration of caBIO objects (e.g. protein sequence, xrefs to mRNA, posttranslational modifications, …)
- Tools
  - Graphic output (GIF or SVG produced by GraphViz) via Java API (as in the current web application)
  - Overlay of expression data via Java API (similar to the current web application)
- External representation -- BioPAX